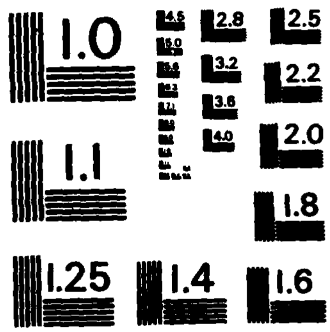
AD-A160 341 AFFINE INVARIANT RANK METHODS IN THE BIVARIATE LOCATION MODEL(U) PENNSYLVANIA STATE UNIV UNIVERSITY PARK DEPT OF STATISTICS B M BROWN ET AL. SEP 85 TR-58 1/1

UNCLASSIFIED N00014-80-C-0741 F/G 12/1 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

12

The Pennsylvania State University

Department of Statistics

University Park, Pennsylvania

TECHNICAL REPORTS AND PREPRINTS

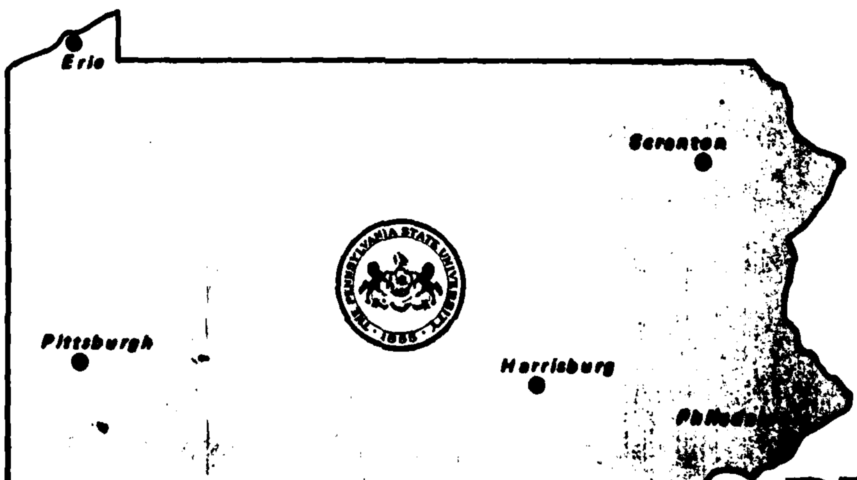
Number 58: September 1985

AFFINE INVARIANT RANK METHODS IN THE BIVARIATE LOCATION MODEL

B. M. Brown
University of Tasmania

and

Thomas P. Hettmansperger*
The Pennsylvania State University



DEPARTMENT OF STATISTICS

The Pennsylvania State University
University Park, PA 16802 U.S.A.

TECHNICAL REPORTS AND PREPRINTS

Number 58: September 1985

AFFINE INVARIANT RANK METHODS IN THE BIVARIATE LOCATION MODEL

B. M. Brown
University of Tasmania

and

Thomas P. Hettmansperger*
The Pennsylvania State University

*The work of this author was partially supported by ONR Contract N00014-80-C0741.

-A-

ABSTRACT

The generalized multivariate median of H. Oja is used to define a multivariate notion of quantile, or rank, and to define a measure of scatter of multivariate linear models. The latter, when applied to the one- and two-sample bivariate location models, yields affine invariant analogs of the Wilcoxon rank-sum and signed-rank tests, and of the corresponding estimates. Additional keywords :

KEY WORDS: affine invariance, bivariate location model, dispersion measures, generalized median, multivariate linear models, multivariate quantile, multivariate rank, permutation tests, R-estimates, Wilcoxon tests.

rank statistics.

1. INTRODUCTION

This paper introduces a notion of multivariate quantile or rank and uses it to develop affine invariant analogs of rank tests and R-estimates in the one- and two-sample bivariate location models.

Bickel (1964) investigates the non-affine invariant vectors of medians and medians of pairwise averages. These are the Hodges-Lehmann (1963) R-estimates derived from the application of the univariate sign and Wilcoxon signed rank statistics, respectively, to the components. In comparing these estimates with the sample mean vector, Bickel concludes that despite encouraging results on robustness and efficiency there remains some pathological behavior of these estimates when the components of the data vectors are highly correlated. He further concludes that the bad behavior may be due in part to the lack of affine invariance of these estimates. Bickel (1965) draws a similar conclusion for tests based on vectors of univariate rank statistics. A different robust estimate, the spatial median $\hat{\theta}$, which minimizes the sum of distances from $\hat{\theta}$ to the data vectors also fails to be affine invariant; see Gower (1974) and Brown (1983). In addition, there may be compelling reasons based on the measurement scales in the model to seek affine invariant rank methods.

Oja (1983) defines a generalized median $\hat{\theta}$ which minimizes a measure of scatter defined by the sum of areas of triangles formed by taking θ along with pairs of data points as vertices. The Oja generalized median is affine invariant. Oja and Niinimaa (1985) study the efficiency of the generalized median and, in the case of bivariate normality, show it to be as efficient as the spatial median.

We introduce a bivariate quantile (or rank) based on the gradient of Oja's measure of scatter. We use this quantile to develop affine invariant tests and estimates in the one- and two-sample bivariate location models. The tests are analogs of the Wilcoxon signed rank test and the Mann-Whitney-Wilcoxon rank sum test, respectively, and the estimates are bivariate analogs of R-estimates. Our approach is similar to that of Jaeckel (1972) and developed in Hettmansperger (1984, Chapter 5). We first construct a measure of dispersion of residuals in a linear model. The dispersion is a linear function of the bivariate residuals in which the coefficients depend on the size or quantile of the residuals. This dispersion, which is related to Oja's scatter measure, provides estimates through minimization and tests from its gradient vector.

The quantiles and generalized median are discussed in Section 2 and the dispersion based on quantiles is defined in Section 3. The two- and one-sample location models are treated in Sections 4 and 5, respectively, and the statistics are illustrated in Section 6.

2. THE BIVARIATE QUANTILE

Let $x_1,\ldots,x_n,\theta$ be 2×1 vectors and let

$$T(\theta) = \sum_{i<j}\sum A(x_i,x_j;\theta) \tag{1}$$

where $A(x_i,x_j;\theta)$ is the area of the triangle formed with x_i,x_j, and θ as vertices. This is the Oja (1983) measure of scatter. The value $\hat{\theta}$ which minimizes $T(\theta)$ is the Oja generalized median of the bivariate sample.

Given $x^T = (x_1, x_2)$, define $\dot{x}^T = (-x_2, x_1)$. Then $\dot{x}$ has the same length as x and is rotated through $\pi/2$ radians in a counter clockwise direction. It follows that

$$(2) \qquad T(\theta) = \tfrac{1}{2} \sum_{i<j}\sum \left|(\dot{x}_j - \dot{x}_i)^T(\theta - x_i)\right|.$$

The quantile vector $Q(\theta)$ is defined by $Q(\theta) = \nabla T(\theta)$, the vector of partial derivatives of $T(\theta)$. Thus the bivariate quantile has both magnitude and direction. Quantiles with largest magnitude correspond to θ being near or beyond the convex hull of the sample. Those with small magnitude correspond to points well embedded in the sample. Further, $-Q(\theta)$ provides the direction of steepest descent on the convex surface defined by $T(\theta)$ and points towards the mass of the sample. An equivalent definition of the Oja generalized median $\hat{\theta}$ is

$$(3) \qquad Q(\hat{\theta}) \doteq 0,$$

where "$\doteq 0$" means that $|Q(\hat{\theta})|$ is a minimum. The equation (3) may determine a single point or a convex set of points from which the median can be selected; see Oja (1983).

Note that

$$(\dot{x}_j - \dot{x}_i)^T(\theta - x_i) = \det\begin{pmatrix} x_i & x_j & \theta \\ 1 & 1 & 1 \end{pmatrix}.$$

Now, from (2), it is easy to show that

$$Q(\theta) = \tfrac{1}{2} \sum_{i<j}\sum u(x_i,x_j;\theta) \tag{4}$$

where

$$u(x_i,x_j;\theta) = \operatorname{sgn}\left\{\det\begin{pmatrix} x_i & x_j & \theta \\ 1 & 1 & 1 \end{pmatrix}\right\}(\dot{x}_j-\dot{x}_i). \tag{5}$$

The vector $u(x_i,x_j;\theta)$ has magnitude $|x_i-x_j|$ and direction perpendicular to and away from the chord defined by (x_i,x_j) toward θ; that is, $u(x_i,x_j;\theta) = \dot{x}_i-\dot{x}_j$ if the order x_i,x_j,θ is clockwise, but $\dot{x}_j-\dot{x}_i$ if the order x_i,x_j,θ is counter clockwise. Hence, $Q(\theta)$ is $\tfrac{1}{2}$ the sum of "repulsion" vectors $u(x_i,x_j;\theta)$ away from the chords defined by pairs of points (x_i,x_j) toward θ.

Using the geometry described in the previous paragraph, or by algebraic manipulation of (4), it follows that

$$\sum Q(x_i) = 0, \tag{6}$$

so the sample quantiles are centered.

In addition, the following observations can be helpful in determining quantiles or locating the generalized median:

(i) When three chords form a triangle, the sum of repulsion vectors is zero for any θ within the triangle; see Figure 1. More generally, if A is a closed, convex loop of chords, the sum of repulsion vectors is zero for any θ in A.

(ii) If θ is on the line extended indefinitely through two data points, then the repulsion vector due to those points is zero.

(iii) If θ lies outside of a triangle of chords, then the sum of repulsion vectors due to the chords is twice the repulsion vector of the most transverse side. See Figure 1. Calculating rules are possible for other convex polygons, but they are too complicated to be of much practical value.

The graph of all lines through pairs of data points (x_i, x_j) and extended indefinitely in both directions breaks the plane into many polygonal regions. The quantile $Q(\theta)$ changes as θ passes from one region to another. The quantile on a border is the average of the quantiles in the contiguous polygons. A computer algorithm is the most efficient way of computing quantiles. The following remarks provide some insight into the computation of quantiles without using a computer.

(a) To find $Q(\theta)$ use (i) to eliminate as many closed loops of chords containing θ as possible. Parts (ii) and (iii) often provide further reductions. Then $Q(\theta)$ is the sum of the repulsion vectors due to the remaining chords.

(b) To locate the generalized median $\hat{\theta}$ (or median set) delete successive closed loops using (i) and beginning with the convex hull and moving inward. When no further reduction is possible, the resulting configuration of extended chords must be examined to find $\hat{\theta}$ that minimizes $|Q(\theta)|$. Generally, all that remains is either one region, whence all θ inside are medians, or one region cut by concurrent diagonals, whence the intersection point of the diagonals is the median. It is quite possible, however, that $\hat{\theta}$ is on a border between polygonal regions. This method is equivalent to deleting polygons in stages; at each stage, delete all regions with a side which is part of the current outer boundary. A new boundary forms at each stage. Stop when there are no further boundaries to eliminate. See Figure 2.

Thus the quantile vector generalizes the idea of a centered rank in a univariate sample. The magnitudes $|Q(x_i)|$ order the depth of the observations in the sample, and the directions $-Q(x_i)$ point toward the center of the data.

In the next section we introduce a measure of dispersion based on the quantiles. We show that it is related to the Oja scatter measure (1).

- Put Figures 1 and 2 about here -

3. DISPERSION

Jaeckel (1972) derived R-estimates in the linear model from a measure of dispersion of the residuals. In the univariate linear model, let the residual r_i be given by $r_i = y_i - \alpha - z_i^T\beta$ where z_i^T is a $1\times p$ row vector of known values, β is a $p\times 1$ vector of unknown regression parameters and α is an unknown scalar intercept parameter. Then an R-estimate of β is defined as the vector $\hat{\beta}$ that minimizes

$$D(\beta) = \sum_{i=1}^{n} [\mathrm{Rank}(y_i - z_i^T\beta) - (n+1)/2](y_i - z_i^T\beta). \tag{7}$$

This dispersion measure is invariant with respect to α. Jaeckel showed that $\hat{\beta}$ generalizes the notion of an R-estimate from simple location models to the linear model. McKean and Hettmansperger (1976) developed tests for

$H\beta = 0$, based on (7), where H is a specified $q \times p$ matrix. See Hettmansperger (1984, Chapter 5) for the details.

The multivariate linear model can be built by appending univariate linear models in the following fashion:

Let Y be an $n \times q$ matrix in which the n rows are independent random vectors such that

$$(8) \qquad EY = E\begin{pmatrix} Y_1^T \\ \vdots \\ Y_n^T \end{pmatrix} = Z\beta$$

where Z is an $n \times p$ matrix of given regression constants and β is a $p \times q$ matrix of unknown parameters. If $Y^{(i)}$ and $\beta^{(i)}$ are the ith columns of Y and β, respectively, then $EY^{(i)} = Z\beta^{(i)}$ is the univariate linear model described in the previous paragraph.

Let $r_i = Y_i - \beta^T z_i$ denote the ith $q \times 1$ residual vector where z_i^T is the ith row of Z. Then (7) becomes

$$(9) \qquad D(\beta) = \sum Q^T(r_i)r_i.$$

In the following sections we will consider the special cases of the two- and one-sample bivariate location models. First, however, we will show that in the bivariate case ($q = 2$), $D(\beta)$ given by (9) is related to Oja's scatter measure (1).

Theorem.

$$D(\beta) = 4 \sum_{i<j<k}\sum\sum A(r_i, r_j, r_k) \tag{10}$$

Proof. Since $\sum Q^T(r_i) = 0$ we have $\sum_i Q^T(r_i) r_j = 0$.

Hence,

$$\begin{aligned} D(\beta) &= \sum_i Q^T(r_i) r_i \\ &= \sum_i Q^T(r_i)(r_i - r_j) \\ &= \sum_i \sum_{j<k}\sum 1/2 \operatorname{sgn}\left\{\det\begin{pmatrix} r_j & r_k & r_i \\ 1 & 1 & 1\end{pmatrix}\right\} (\dot{r}_k - \dot{r}_j)^T (r_i - r_j) \\ &= \sum_i \sum_{j<k}\sum A(r_j, r_k, r_i) \\ &= 1/2 \sum_i \sum_j \sum_k A(r_j, r_k, r_i) \\ &= 1/2 . 8 . \sum_{i<j<k}\sum\sum A(r_j, r_k, r_i) \end{aligned}$$

Thus, our dispersion measure is proportional to the sum of areas of triangles with residuals at the vertices. The scale invariant R-like estimate is the matrix $\hat{\beta}$ that minimizes this sum of triangular areas.

4. THE TWO SAMPLE LOCATION MODEL

In the bivariate two-sample location model the matrix β in (8) can be replaced by a vector. The intercept part of the linear model does not affect the difference in locations. Accordingly, in the bivariate two-sample problem there are n_1 observations $x_1, \ldots, x_{n_1}$ and n_2 observations $y_1, \ldots, y_{n_2}$. Let Δ be the location shift vector applied to the y- sample, so that the residuals are either x_i or $y_j-\Delta$. Then the terms of D are areas of triangles whose vertices number s from the x-sample and 3-s from the y- sample, for s = 0,1,2,3. The next result shows that, similar to the univariate case, the dispersion depends on the y-x differences.

Theorem.

(11)
$$D(\Delta) = 4\Big\{\text{constant} + \sum_k \sum_{i<j} A(y_k-x_i, y_k-x_j, \Delta) + \sum_i \sum_{j<k} A(y_j-x_i, y_k-x_i, \Delta)\Big\}$$

Proof. Apply D in (10) to the combined samples. Note that areas are not affected by the same displacement applied to all three vertices or by sign changes. Hence, areas that involve three x's or three y's do not depend on Δ. We now have

$$D(\Delta) = 4\Big\{\text{constant} + \sum_k \sum_{i<j} A(x_i, x_j, y_k-\Delta) + \sum_i \sum_{j<k} A(y_j-\Delta, y_k-\Delta, x_i)\Big\},$$

but $A(x_i,x_j,y_k-\Delta) = A(y_k-x_i,y_k-x_j,\Delta)$ and $A(y_j-\Delta,y_k-\Delta,x_i) = A(y_j-x_i,y_k-x_i,\Delta)$. This completes the argument.

The gradient of $D(\Delta)$ is given by

(12)
$$S(\Delta) = 1/2 \sum_k \sum_{i<j} u(y_k-x_i, y_k-x_j; \Delta) + 1/2 \sum_i \sum_{j<k} u(y_j-x_i, y_k-x_i; \Delta).$$

It is sufficient to consider testing the null hypothesis H_0: $\Delta = 0$ against either a general alternative H_1: $\Delta \neq 0$ or some directional alternative stating that Δ differs from 0 in some fixed direction. A test based on the quantiles uses $S = S(0)$, the gradient of $D(\Delta)$ evaluated at the hypothesized value.

The next result shows that, like the univariate rank sum statistic, S reduces to the sum of the x-quantiles in the combined sample.

<u>Theorem.</u>

(13)
$$S = \sum_{i=1}^{n_1} Q(x_i)$$

<u>Proof.</u> From (12) we have

$$2S = \sum_k \sum_{i<j} u(-x_i,-x_j;-y_k) + \sum_i \sum_{j<k} u(y_j,y_k;x_i)$$

$$= \sum_i \sum_{j<k} u(y_j,y_k;x_i) - \sum_k \sum_{i<j} u(x_i,x_j;y_k).$$

Note that $u(x_i,x_j;y_k) + u(x_j,y_k;x_i) + u(y_k,x_i;x_j) = 0$ and recall from (6) that $\sum_i \sum_{j<k} u(x_j,x_k;x_i) = 0$. With these facts, 2S reduces to

$$\sum_i \sum_{j<k} u(y_j,y_k;x_i) + \sum_{i \neq j} \sum_k u(x_j,y_k;x_i)$$

$$+ \sum_i \sum_{j<k} u(x_j,x_k;x_i)$$

which, when compared to (4), is seen to be the result stated in the theorem.

The test statistic is a clear analog of the Mann-Whitney-Wilcoxon rank sum statistic in the univariate case. The set $x_1 + \Delta, \ldots x_{n_1} + \Delta$ is one of $\binom{n_1+n_2}{n_1}$ equally likely subsamples of n_1 bivariate observations drawn from the combined $x + \Delta$ and y set of size $n_1 + n_2$. Hence, standard permutation arguments show that $S(\Delta) = \sum_i Q(x_i+\Delta)$ has expectation 0. The natural estimate of Δ is $\hat{\Delta}$ such that $S(\hat{\Delta}) \doteq 0$. This is the analog of the Hodges-Lehmann (1963) estimate of shift in the univariate two sample problem. Equation (12) shows that $\hat{\Delta}$ is an Oja generalized median computed on the cross sample differences.

Under the null hypothesis H_0: $\Delta = 0$, the permutation argument shows that $ES = 0$ and the covariance matrix of S is

$$C = \frac{n_1 n_2}{(n_1+n_2)(n_1+n_2-1)} \sum_{i=1}^{n_1+n_2} Q(z_i)Q^T(z_i) \tag{14}$$

where $z_1,\ldots,z_{n_1+n_2}$ represents the combined sample. Furthermore, S will be approximately bivariate normal for large n_1,n_2.

An approximately size α test for H_0: $\Delta = 0$ against H_A: $\Delta \neq 0$ rejects H_0 if $S^TC^{-1}S > \chi^2_\alpha(2)$ where $\chi^2_\alpha(2)$ is the $1-\alpha$ percentile from a chi-square distribution with 2 degrees of freedom. To test H_0: $\Delta = 0$ against an alternative specifying a fixed direction with unit vector v, project on v yielding v^TS with null covariance matrix v^TCv. We reject H_0 if $v^TS/(v^TCv)^{\frac{1}{2}} > z_\alpha$ where z_α is the $1-\alpha$ percentile from the standard normal distribution.

5. THE ONE SAMPLE LOCATION MODEL

Suppose $x_1,\ldots,x_n$ is a sample from a bivariate distribution with the property that $x-\theta$ and $\theta-x$ have the same distribution. Then θ represents the center of the distribution.

In testing H_0: $\theta = 0$, it is difficult to develop a simple sign-test analog based on the Oja generalized median $\hat{\theta}$, (6). The natural test function is $Q = Q(0) = \frac{1}{2} \sum_{i<j} \sum u(x_i, x_j; 0)$, but a simple randomization argument does not provide the null distribution. The main goal of this section is to develop an analog of the Wilcoxon signed rank test.

A standard device for producing one-sample univariate rank methods from two-sample procedures is to create a second, artificial sample that consists of the negatives of the original sample. When we consider the univariate rank of $-x_1$, say, relative to $x_1, \ldots, x_n$ the result is the number of sums $x_1+x_j, j=1,\ldots,n$ [or averages $(x_1+x_j)/2$] that are negative. By doing this for each data point, we find that the two-sample rank method produces counts of the signs of the pairwise sums or averages, and these counts are related to the ranks of the absolute values; see Hettmansperger (1984, Section 2.3). Hence, we arrive at the one sample signed rank statistics. This device, in the bivariate case, allows us to avoid the problem of introducing absolute values in the plane.

Returning to the bivariate case, let $-x_1, \ldots, -x_n$ be the second, artificial sample. For inference on θ, following the discussion in the previous section, let $\Delta = 2\theta$ and consider

$$S(\Delta) = \sum_{i=1}^{n} Q_{2n}(-x_i+\Delta) \tag{15}$$

where the subscript 2n on Q indicates that the quantile of $-x_i+\Delta$ is computed relative to $-x_1+\Delta, \ldots, -x_n+\Delta, x_1, \ldots, x_n$. The next result shows that we need only consider $Q_n(-x_i+\Delta)$; that is, the quantile of $-x_i+\Delta$ relative to $x_1, \ldots, x_n$.

<u>Theorem.</u>

$$S(\Delta) = 2 \sum_{i=1}^{n} Q_n(-x_i+\Delta). \tag{16}$$

Proof. Let $\Delta = 0$ without loss of generality. Now

$$4Q_{2n}(-x_i) = \sum_j \sum_k u(x_j, x_k; -x_i)$$

$$+ \sum_j \sum_k u(-x_j, x_k; -x_i) + \sum_j \sum_k u(x_j, -x_k; -x_i)$$

$$+ \sum_j \sum_k u(-x_j, -x_k; -x_i)$$

$$= A_i + B_i + C_i.$$

Summing on i, $\sum C_i = 0$. Since

$u(a,b;c) = -u(b,c;a) - u(a,c;b)$ and $u(a,b;c) = -u(-a,-b;-c)$,

$$\sum B_i = \sum_i \sum_j \sum_k \{-u(x_k, -x_i; -x_j) - u(-x_j, -x_i; x_k)$$

$$- u(-x_k, -x_i; x_j) - u(x_j, -x_i; -x_k)\}$$

$$= \sum_i \sum_j \sum_k \{-u(x_k, -x_i; -x_j) + u(x_j, x_i; -x_k)$$

$$+ u(x_k, x_i; x_j) - u(x_j, -x_i; -x_k)\}.$$

Group the first with the fourth terms and the second with the third terms to get $\sum B_i = -\sum B_i + 2\sum A_i$. Hence, $\sum B_i = \sum A_i$ and

$$4\sum Q_{2n}(-x_i) = 2\sum A_i$$

$$= 2.2 \sum_i \sum_{j<k} \sum u(x_j, x_k; -x_i)$$

$$= 8 \sum 1/2 \sum_{j<k} \sum u(x_j, x_k; -x_i)$$

which reduces with an application of (4) to the result stated in the theorem.

This theorem shows that the estimate of location $\hat{\theta} = \hat{\Delta}/2$ is defined by $S(\Delta) \doteq 0$. A further interpretation is possible. Note that

$$Q_n(-x_i+\Delta) = \tfrac{1}{2} \sum_{j<k}\sum u(x_j, x_k; -x_i+\Delta)$$

$$= \tfrac{1}{2} \sum_{j<k}\sum u(x_i+x_j, x_i+x_k; \Delta).$$

Hence,

$$S(\Delta) = \tfrac{1}{2} \sum_{j<k}\sum \sum_{i} u(x_i+x_j, x_i+x_k; \Delta),$$

or

(17) $$S(\theta) = \tfrac{1}{2} \sum_{j<k}\sum \sum_{i} u\left(\frac{x_i+x_j}{2}, \frac{x_i+x_k}{2}; \theta\right).$$

Analogous to the univariate Hodges-Lehmann estimate which is the median of the pairwise averages, the bivariate estimate $\hat{\theta}$ is an Oja generalized median computed on the coupled pairs displayed in (17). In effect, (17) shows how the data can be symmetrized before the median operation is applied.

For testing the hypothesis H_0: $\theta = 0$ we could use either $\sum Q_{2n}(-x_i)$ or $\sum Q_n(-x_i)$. Under H_0: $\theta = 0$ the randomization distribution of $\sum Q_{2n}(-x_i)$ is easier to work with. The first line of the proof of (16) shows that $Q_{2n}(-x_i) = -Q_{2n}(x_i)$. Under the null hypothesis each has probability 1/2 so that the statistic $S = \sum_i Q_{2n}(-x_i)$ has $ES = 0$ and covariance matrix

$$C = \sum Q_{2n}(-x_i) Q_{2n}^T(x_i).$$

Tests of H_0: $\theta = 0$ are carried out as described in the last paragraph of the last section. The test statistic is a scale invariant bivariate analog of the Wilcoxon signed rank statistic.

6. EXAMPLES

In this Section the two-sample and one-sample bivariate rank methods are illustrated through application to two data-sets. First consider the two sample test. In the following data from Hettmansperger (1984, p.291), the data consist of levels of two biochemical components in brains of mice, $x^T = (X_1, X_2)$.

Control group

x_1:	1.21	0.92	0.80	0.85	0.98	1.15	1.10	1.02	1.18	1.09
x_2:	0.61	0.43	0.35	0.48	0.42	0.52	0.50	0.53	0.45	0.40

Treatment group

y_1:	1.40	1.17	1.23	1.19	1.38	1.17	1.31	1.30	1.22	1.00	1.12	1.09
y_2:	0.50	0.39	0.44	0.37	0.42	0.45	0.41	0.47	0.29	0.30	0.27	0.35

The corresponding quantiles of the control group observations among the combined control plus treatment sample are

x_1:	.89	-9.01	-8.82	-9.40	-6.78	-.51	-3.25	-6.24	2.28	-3.04
x_2:	18.53	3.65	-6.26	9.37	.76	15.70	13.63	15.85	5.27	-4.56

and the sum of quantiles is $S^T = (-43.88, 71.94)$. The standardized test of no location shift between control and treatment populations is $S^T C^{-1} S = 15.137$, which is highly significant when referred to χ_2^2. By comparison, a univariate rank method applied componentwise yields $\chi_2^2 = 14.22$; see Hettmansperger (1984,p.292).

For a one-sample test, the following data from Hettmansperger (1984, p.286) are systolic and diastolic blood pressures of 15 adult male Peruvian Indians.

x_1	170	125	148	140	106	108	124	134	116	114	118	138	134	124	114
x_2	76	75	120	78	72	62	70	64	76	74	68	78	86	64	66

To test that the center of the bivariate population is (120,80), consider the sample of $(y_1,y_2) = (x_1-120,\ x_2-80)$ values and the reflected artificial sample of values $(-y_1,\ -y_2)$. The quantiles of the reflected sample among the combined sample are

y_1	-3271	-1441	-1205	-2814	2064	944	-1346	-2241	596
y_2	1145	1579	-3184	1190	1321	2858	2522	3276	602

y_1	1052.5	-410	-2483	-2298	-1384.5	220
y_2	1088.5	2822	1238	-552	3634.5	2911

and the quantile sum is $S^T = (-14017,\ 22469)$. The standardized statistic $S^TC^{-1}S = 8.49$, highly significant when referred to χ_2^2. The corresponding χ_2^2 for the componentwise univariate rank method has the same value; see Hettmansperger (1984,p.287).

References

Bickel, P.J. (1964). On Some Alternative Estimates of Shift in the P-Variate One Sample Problem, Ann. Math. Statist., 35, 1079-1090.

Bickel, P.J. (1965). On Some Asymptotically Nonparametric Competitors of Hotelling's T^2, Ann. Math. Statist., 36, 160-173.

Brown, B.M. (1983). Statistical uses of the spatial median. J. Roy.

Statist. Soc. B 45, 25-30.

Gower, J.C. (1974). The mediancentre. Appl. Statist. 23, 466-470.

Hettmansperger, T.P. (1984). Statistical Inference based on Ranks. New York: Wiley.

Hodges, J.L. Jr. and Lehmann, E.L. (1963). Estimates of Location Based on Rank Tests, Ann. Math. Statist., 34, 598-611.

Jaeckel, L. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. Ann. Math. Statist., 40, 1889-1900.

Oja, H. (1983). Descriptive Statistics for Multivariate Distributions, Stat. and Prob. Letters, 1, 327-332.

Oja, H. and Niinimaa, A. (1985) Asymptotic properties of the generalized median in the case of multivariate normality. J. Roy. Statist. Soc. 47 (to appear).

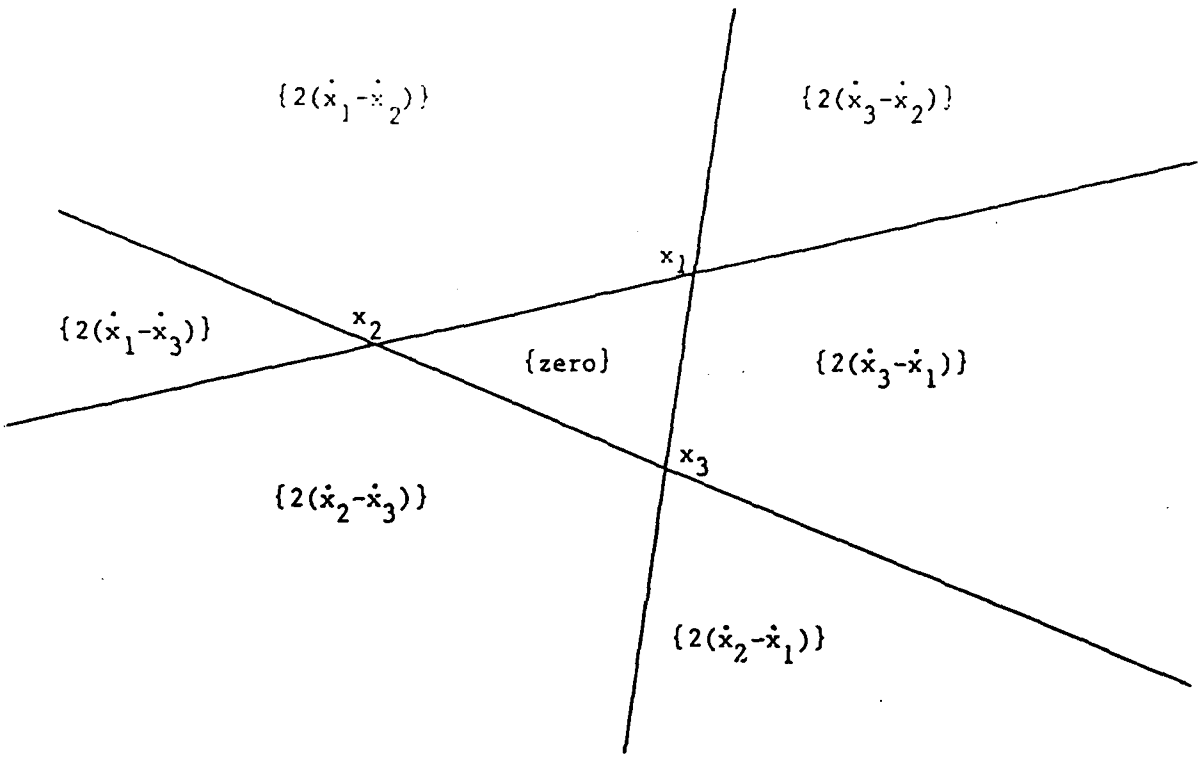


Fig. 1: Sum of repulsion vectors due to the three sides of a triangle (x_1,x_2,x_3) shown in brackets { } for each of seven regions.

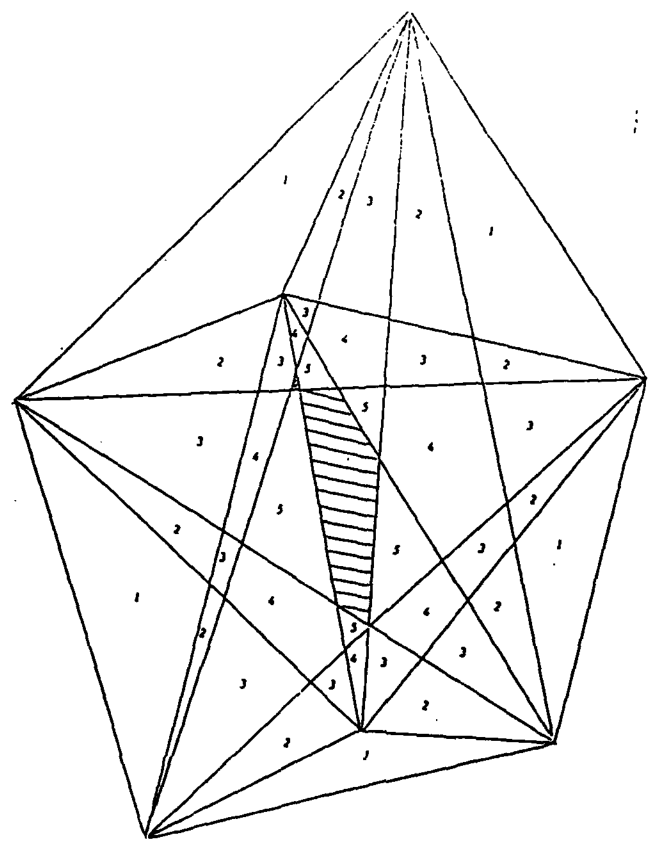


Fig. 2: Median region of seven points is shaded. Regions surrounding it are deleted in order, denoted by 1,2,... (some regions are too small to be numbered).

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 58	2. GOVT ACCESSION NO. AD-A160	3. RECIPIENT'S CATALOG NUMBER 341
4. TITLE (and Subtitle) Affine Invariant Rank Methods in the Bivariate Location Model		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) B. M. Brown, University of Tasmania Thomas P. Hettmansperger, The Pennsylvania State University		8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0741
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics The Pennsylvania State University University Park, PA 16802		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR042-446
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistical and Probability Program Code 436 Arlington, VA 22217		12. REPORT DATE September 1985
		13. NUMBER OF PAGES 17
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Affine invariance, bivariate location model, dispersion measures, generalized median, multivariate linear models, multivariate quantile, multivariate rank, permutation tests, R-estimates, Wilcoxon tests.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The generalized multivariate median of H. Oja is used to define a multivariate notion of quantile, or rank, and to define a measure of scatter of multivariate linear models. The latter, when applied to the one- and two-sample bivariate location models, yields affine invariant analogs of the Wilcoxon rank-sum and signed-rank tests, and of the corresponding estimates.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)